NBER WORKING PAPER SERIES

REVISITING SOME PRODUCTIVITY DEBATES

Johannes Van Biesebroeck

Working Paper 10065
http://www.nber.org/papers/w10065

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2003

I would like to thank Jim Tybout and Mark Roberts for providing me with the data and Mel Fuss, Frank Wolak, Marc Melitz, and participants at the NBER Summer Institute 2002 and the summer workshop 2003 at the CEU for comments. All remaining errors are my own. Financial support from the Connaught Fund is gratefully acknowledged. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

Revisiting Some Productivity Debates
Johannes Van Biesebroeck
NBER Working Paper No. 10065
October 2003
JEL No. C3, O3, O4

ABSTRACT

Researchers interested in estimating productivity can choose from an array of methodologies, each with its strengths and weaknesses. Methods differ by the assumptions they rely on and imply very different calculations. I compare five widely used techniques: (a) index numbers, (b) data envelopment analysis, and three parametric methods, (c) instrumental variables estimation, (d) stochastic frontiers, and (e) semi-parametric estimation. I compare the estimates directly and evaluate three productivity debates using a panel of manufacturing plants in Colombia. The different methods generate surprisingly similar results. Correlations between alternative productivity estimates are invariably high. All methods confirm that exporters are more productive on average and that only a small portion of the productivity advantage is due to scale economies. Productivity growth is correlated more strongly with export status, frequent investments in capital equipment, and employment of managers than with the use of imported inputs or foreign ownership. On the debate whether aggregate productivity growth is driven by plant-level changes or output share relocation, all methods point to the importance of plant-level changes, in contrast to results from the U.S.

Johannes Van Biesebroeck
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON M5S 3G7
CANADA
and NBER
jovb@chass.utoronto.ca

1 Motivation

Productivity is used and discussed widely. Ever since Solow (1957) decomposed output growth into the contribution of input growth and a residual productivity term, the concept has increased in popularity. Productivity has generated a lot of interest in its own right and is used as a benchmark to rank firms or countries. Such rankings gained credibility once other studies documented that productivity is correlated with other indicators of success such as employment growth, export status, or technology adoption. Low productivity has also been found to predict exit, the ultimate performance standard. Its importance can also be gauged from the attention it receives as a criterion to evaluate policy interventions or firms' decisions. In industrial economics, for example, a large literature investigates the effect of R&D on productivity and the resulting impact on industry structure. In international economics, efforts to evaluate the impact of trade liberalization has turned from estimating changes in price-cost margins to productivity changes.

Fundamentally, the objective of productivity measurement is to identify output differences that cannot be explained by input differences. Because the production technology of each firm and the input tradeoff it allows, is not observed, our ability to control for input substitution is subject to error. In addition, inputs and outputs are likely to be measured with error, certainly in less intensively used data sets from developing countries. Methodologies for productivity measurement differ vastly in their sensitivity to measurement and specification error.[1]

I evaluate five widely-used methodologies, which fall in three broad classes. The first two, index numbers and data envelopment analysis, are flexible in the specification of technology, but do not allow for measurement errors in the data. The other three parametric methods calculate productivity from an estimated production function. Because the framework is explicitly stochastic they are less vulnerable to measurement error, certainly in the dependent variable, but misspecification of the production function might be an issue. I use three estimators for the production function that address the simultaneity of productivity and input choices differently: instrumental variables, stochastic frontiers, semiparametric estimation.

[1] See Van Biesebroeck (2003c) for a comparison in a Monte Carlo framework.

Using a sample of manufacturing plants from Colombia, I compare productivity estimates with the different methods directly and I review a number of debates with important policy implications. The results indicate that the productivity level and growth estimates are surprisingly similar. Partial correlation coefficients across most methodologies are invariably high. From a policy perspective, the important issue is whether the different methods reach the same conclusions on underlying economic phenomena. In particular, I find that the answer to the following three questions is robust to measurement methodology.

- Are exporters more productive than nonexporters?
- What causes technological change? Better capital equipment, foreign expertise, or high levels of human capital?
- Is aggregate productivity growth driven by plant-level productivity increases or by the reallocation of inputs from less to more productive establishments?

The three debates have received a fair amount of attention in the literature on economic development. Because they involve very different aspects of the productivity distribution—the first question compares productivity levels across firms, the second compares growth rates, and the third question relies on changes in the entire productivity distribution—they allow a comprehensive overview of the impact of measurement methodology.

In the next section, I give some background on productivity measurement and introduce the different methodologies. Only the general idea and crucial equations are presented to convey the distinctive features of each methodology. Links to the literature for more detailed information are provided in the respective sections. Section 3 introduces the data and directly compares productivity measures across methodologies. Subsequently, I verify whether the answers to each of the three debates vary by estimation methodology. The final section summarizes what the comparisons teach us about the measuring methodologies and about the economic phenomena they describe.

2 Measuring Productivity

In plain English, one firm is more productive than another if it is able to produce the same outputs with less inputs or if it produces more outputs using the same inputs. Similarly, a firm has experienced positive productivity growth if outputs have increased more than inputs or inputs decreased more than outputs. The comparison is more difficult when a first production plan uses more of one input, while a second plan relies more on a second input, keeping output constant. Clearly, the possibility for input substitution embedded in the technology will govern such a comparison. This makes it impossible to talk about productivity without specifying a production function (or a cost function or any other representation of technology). Measuring productivity necessarily involves decomposing differences in the input-output combination into shifts along a production function and shifts of the function itself.

Figure 1 illustrates the objective of productivity comparisons. In the simple case of one output, which will be maintained throughout, it compares two unit production plans, P_0 and P_1, in input space. Part of the difference, from P_0 to 1, is a shift along the frontier —represented here by the unit isoquant— exploiting the input substitution possibilities of the technology. Part of the difference, from 1 to P_1, is an actual shift of the frontier, which is counted as technical change or productivity growth. If technical change is restricted to be Hicks-neutral, the shape of the unit isoquant will be unchanged after the shift. Capital-biased or labor-saving technical change, on the other hand, makes the isoquant shift more down than left.

[Figure 1]

The shape of the unit isoquant in Figure 1 can be estimated if one is willing to make functional form assumptions. An econometric problem arises as input choices, the explanatory variables in the production function equation, depend on unobserved productivity differences. I will rely on three different solutions to this simultaneity problem. The difficulty is to estimate the shape of the isoquant in Figure 1, which is common across firms, when the observed position along the isoquant of each firm depends on the exact location of the isoquant, which varies across firms.

Alternatively, if factor prices for each firm are observed, one can rely on the theory of index numbers. No production function has to be specified or estimated. Assuming that the first order conditions for profit maximization hold, the observed production plans are tangency points of the budget constraint and isoquant. Assuming further that the isoquants for all firms have the same curvature allows us to infer the relative position of the isoquants.

A third, nonparametric approach constructs a piece-wise linear isoquant to maximize the productivity for the unit under consideration, with the constraint that no other production plan lies below the isoquant. The section of the isoquant that intersects the line through the origin for the observation under consideration, implicitly defines relative weights for labor and capital. The weights are calculated separately for each observation using linear programming techniques to minimize the distance to the isoquant.

Each method compares two production plans, it and $j\tau$, which can refer to two different firms ($t = \tau$) or to the production plans of a single firm at two different points in time ($i = j$). I adopt a number of standard assumptions—single output[2], only Hicks-neutral productivity differences[3], output-based productivity comparisons[4]—and abstract from a number of thorny issues that researchers have dealt with.[5] The production function

$$Q_{it} \;=\; A_{it}\; F_{(i)}(X_{it}),$$

concentrates all productivity differences in the multiplicative factor A_{it}, which differs between

[2]Using deflated sales or value added as a single output implies aggregation of different products using firm-specific prices.

[3]In practice, most studies use a Cobb-Douglas production function, which makes it impossible to identify the factor-bias of technological change.

[4]The output-based comparison provides an answer to the question: "How much extra output does a firm produce, relative to another firm, conditional on its (extra) input use?" An alternative measure is input-based and asks "What is the minimum input requirement for one firm to produce the same output as another firm?" Under constant returns to scale the two measures coincide for each of the methods I adopt.

[5]If competition is less than perfect, the use of deflated sales as output measure is problematic. Klette and Griliches (1996) provide a solution if one is willing to make assumptions on the type of competition and functional form of demand. In the first study using a full census of manufacturing plants, Griliches and Ringstad (1971) discuss the relative merits of a value added and gross output production function. The problems associated with the aggregation of inputs and outputs, as well as quality differences in heterogeneous inputs are the subject of an exchange between some of the pioneers of productivity decompositions, see Jorgenson and Griliches (1967) and Denison (1972). Methods have also been developed to deal with variations in capacity utilization (Berndt and Fuss (1986)) and regulated firms (Denny et al. (1981)).

firms and changes over time. If the production function $F(.)$ varies between firms it has to be specified which firm's technology (i or j) is used for productivity comparisons.

In this framework, productivity comparisons boil down to the ratio of ratios in equation (1), which illustrates that productivity is intrinsically a relative concept:

$$\frac{A_{it}}{A_{j\tau}} = \frac{Q_{it} \big/ Q_{j\tau}}{F(X_{it}) \big/ F(X_{j\tau})}. \tag{1}$$

The calculation of the denominator in (1)—the ratio of input aggregators—distinguishes the different methods.[6]

Readers familiar with the different methodologies might still find the following expositions useful, as it summarizes a number of different literatures in a unified framework.[7]

2.1 Data Envelopment Analysis (DEA)

The first approach to productivity measurement is completely nonparametric and uses linear programming. The method dates back to Farrell (1957) and it was operationalized by Charnes, Cooper, and Rhodes (1978).[8] No particular production function is assumed. Instead, the ratio of a linear combination of outputs over a linear combination of inputs is compared across observations. The intuition is to lay a piece-wise linear production frontier in input-output space over the most efficient observations, as in Figure 2. Observations that are not dominated are labeled 100% efficient. Domination occurs when another firm, or a linear combination of

[6]For level comparisons, the benchmark in the denominator will be the average productivity level for all plants in the industry, $\overline{A_t}$, which facilitates multilateral comparisons. In practice, most studies have used $\log A_{it} - \overline{\log A_t}$ as multilateral productivity comparison, taking the average of the logarithm. For comparability, I follow this practice.

[7]I try to avoid the frontier versus nonfrontier distinction. Some authors argue that some output shortfall, given inputs, is the result of inefficiency at the firm-level. To be consistent with a profit maximizing, I include such shortfalls in productivity as they might be caused by technology differences, unmeasured inputs, or quality differences in outputs, among other possibilities. Rather than placing some firms below the production frontier, I assume that they are on their own *firm-specific* frontier which lies below the industry best-practice. The difference represents lower productivity. See Stigler (1976) for a more elaborate and more powerful motivation. An alternative perspective is provided in Coelli, Rao, and Battese (1997), which reviews the index, DEA, and stochastic frontier methods with the goal of efficiency measurement.

[8]More information on the method and applications can be found in Seiford and Thrall (1990).

other firms, uses less of all inputs to produce the same outputs (for an input-based measure) or produces more of all outputs using the same inputs (for an output-based measure). Multiple inputs or outputs are aggregated linearly with weights chosen optimally for the unit under consideration. Weights are restricted such that the efficiency of all observations does not exceed 100% when the same weights are applied to them.

The example in Figure 2 is drawn for a single input and output, but the intuition is similar for higher dimensional problems as inputs and outputs are aggregated linearly. P_1 to P_5 represent the production plans of different firms. The solid line represents the frontier if variable returns are allowed. Four of the five observations lie on the frontier and are deemed 100% efficient. If the technology is restricted to constant returns to scale, the frontier is forced to go through the origin and is extrapolated beyond observed data points, resulting in the dashed line as production frontier. Only the second plan is efficient in this case.[9] The distance of each unit to the frontier measures the estimated efficiency. In an input orientation, one improves efficiency by reducing inputs: a horizontal projecting onto the frontier. In an output orientation, efficiency is increased by increasing output until the unit produces on the frontier given its observed inputs: a vertical projection.[10]

[Figure 2]

A linear programming problem is solved separately for each observation. Input and output weights are chosen to maximize efficiency (θ_1). The number of restrictions equals the number of observations, plus sign restrictions on the weights. For unit 1 the problem amounts to

$$\begin{aligned} \max_{v_l, u_k} \quad & \theta_1 = \frac{\sum_{l=1}^{L} v_l q_1^l}{\sum_{k=1}^{K} u_k x_1^k} \\ \text{subject to} \quad & \frac{\sum_l v_l q_i^l}{\sum_k u_k x_i^k} \leq 1 \qquad i = 1...N \\ & v_j,\ u_k \geq 0 \qquad l = 1...L,\ k = 1...K, \end{aligned} \tag{2}$$

i indexes firms, l outputs, and k inputs. It is converted to a linear programming problem by

[9] Imposing constant returns to scale adds a constraint to the problem and the maximized objective value will be (weakly) lower.

[10] Clearly, under variable returns both orientations yield different results, as the frontier does not go through the origin and the slope of the segments the unit gets projected onto might differ.

multiplying the objective and restrictions by their denominator and adding a normalization.[11] In practice, most applications solve the dual problem, where θ_1 is chosen directly. The current formulation implicitly incorporate a constant returns to scale assumption. To relax this, an extra slack-variable is introduced in (2) or an extra constraint is added to the dual problem.

The efficiency measures θ_i can be interpreted as the productivity difference between unit i and the most productive unit: $\theta_i = \dfrac{A_i}{A_{\max}}$. To obtain a measure comparable with other methodologies, I define

$$\begin{aligned} \log A_{it}^{DEA} - \overline{\log A}_t^{DEA} &= \log \theta_i - \frac{1}{N}\sum_{i=1}^{N} \log \theta_i, \\ \log A_{it}^{DEA} - \log A_{it-1}^{DEA} &= \log \theta_{it} - \log \theta_{it-1}. \end{aligned} \tag{3}$$

as the relative productivity level and growth rate. Productivity growth is less commonly measured in the DEA framework. Nevertheless, including the different firm-years as separate observations in the analysis, it is possible to calculate productivity growth as indicated. While these transformations are arbitrary, they do not change the ranking of firms, only the absolute productivity levels and growth rates.

DEA has the advantage to deal with many outputs in a consistent way. It leaves the underlying technology unspecified and allows for heterogeneity, without functional form or behavioral assumptions. While there is no theoretical justification for the linear aggregation, it is natural in an activities analysis framework. The flexibility in weighting can be a drawback. It has the implication that each firm with the highest output-input ratio for any combination of outputs and inputs will be considered efficient. The method is not stochastic, which is demanding on the data and makes the method sensitive to outliers.[12] One might object to the label "100% efficient" for the best practice firms in the sample. In some situations no firm might be efficient, e.g. due to regulation.

---

[11] The scale of weights is not defined: multiplying all weights by the same multiplier does not change the problem. Usually, the linear combination of inputs is set to unity for the unit under investigation. Interchanging the roles of inputs and output in (2) and minimizing the objective function, gives the corresponding output-oriented programming problem for $\frac{1}{\theta}$

[12] With variable returns to scale, each firm with the lowest input or highest output level in absolute terms is also fully efficient. More recently, stochastic DEA methods have been developed, but they are not universally accepted yet. Most application still apply the deterministic variants.

2.2 Index numbers (TFP)

The second approach provides a theoretically motivated aggregation method for inputs and outputs, while remaining fairly agnostic on the shape of the underlying technology. Under a number of assumptions, it is possible to calculate the last term in (1) from observables, without having to specify the exact production function, nor forcing it to be uniform across observations.

The work of Solow (1957) and Diewert (1976) led to the total factor productivity (TFP) formula

$$\log \frac{A_{it}}{A_{j\tau}} = \log \frac{Q_{it}}{Q_{j\tau}} - (\tfrac{s^L_{it}+s^L_{j\tau}}{2}) \log \frac{L_{it}}{L_{j\tau}} - (1 - \tfrac{s^L_{it}+s^L_{j\tau}}{2}) \log \frac{K_{it}}{K_{j\tau}}. \tag{4}$$

where s^L_{it} is the fraction of the wage bill in output or total cost. Deducting input differentials weighted by their share in output from the output differential produces an exact measure for Hicks-neutral technical change. At least if a number of assumptions are satisfied: returns to scale are constant, firms are profit maximizing and operate in competitive input and output markets, and the underlying production function is translog.[13]

Caves, Christensen, and Diewert (1982a) extended (4), allowing for technical change that is not Hicks-neutral and variable returns to scale, and giving it a more general interpretation. They start from the Malmquist productivity index and represent the technology by output and input distance functions. Under the same assumptions as before, the geometric mean of the two output-based productivity indices based on each firm's technology, $m^O(x_{it}, x_{j\tau}, y_{it}, y_{j\tau})$, *exactly* equals the difference between a Törnqvist output index and the corresponding input index with a scale factor to account for non-constant returns to scale:[14]

$$\begin{aligned}\log m^O(x_{it}, x_{j\tau}, q_{it}, q_{j\tau}) &= \sum_l^L \tfrac{r^l_{it}+r^l_{j\tau}}{2}(\log q^l_{it} - \log q^l_{j\tau}) - \sum_k^K \tfrac{s^k_{it}+s^k_{j\tau}}{2}(\log x^k_{it} - \log x^k_{j\tau}) \\ &\quad + \sum_k^K \tfrac{s^k_{it}(1-\epsilon_{it})+s^k_{j\tau}(1-\epsilon_{j\tau})}{2} (\log x^k_{it} - \log x^k_{j\tau}).\end{aligned} \tag{5}$$

[13] In the single-output case, only cost minimization is needed. With multiple outputs, the output ratio is replaced by the ratio of a revenue-weighted sum of outputs, similar to the cost shares as input-weights. Only the second order terms in the technology have to be equal for the two units compared.

[14] The geometric mean of the two input-based productivity indices differs only in the scale factor.

r_z^l is the revenue share of output l and firm z ($l = 1...L, z = it, j\tau$), s_z^k is the cost share of input k and firm z ($n = 1...N$), and ϵ^z are the (local) returns to scale for firm z.

In most applications, the scale adjustment is omitted. Only the first two terms of (5) are included, reproducing equation (4), which amounts to lumping the effect of scale economies in the productivity measure. For comparability with the other methodologies, I do include the scale factor.[15] If some conditions do not hold the index number is not exact, but still a valid second-order approximation to the productivity ratio.[16]

Equation (5) accommodates productivity growth and level comparisons. For comparisons between different firms, multilateral comparisons are generally preferred over bilateral ones, because Törnqvist indices are not transitive. Caves, Christensen, and Diewert (1982b) propose an alternative formula, where each firm is compared to a hypothetical firm (with average log-output $\overline{\log Q}$, labor share $\overline{s}^L$, etc.). Assuming returns to scale are equal across observations, the index formulas are

$$\begin{aligned} \log A_{it}^{IN} - \overline{\log A}_t^{IN} &= (\log Q_{it} - \overline{\log Q}_t) - \overline{\epsilon}\Big[\overline{\overline{s}}_{it}(\log L_{it} - \overline{\log L}_t) - (1 - \overline{\overline{s}}_{it})(\log K_{it} - \overline{\log K}_t)\Big] \\ \log A_{it}^{IN} - \log A_{it-1}^{IN} &= \log(Q_{it}/Q_{it-1}) \;-\; \overline{\epsilon}\;\Big[\overline{s}_{it}\;\log(L_{it}/L_{it-1}) \;-\; (1 - \overline{s}_{it})\;\log(K_{it}/K_{it-1})\Big] \end{aligned} \quad (6)$$

where $\overline{s}_{it} = \frac{s_{it}^L + s_{it-1}^L}{2}$ and $\overline{\overline{s}}_{it} = \frac{s_{it}^L + \overline{s}_t^L}{2}$. This permits multilateral comparisons, yields bilateral comparisons that are transitive, and still allows for technology that is firm-specific.

One of the main advantages of the index number approach is the ease of calculation. Also, the specification of technology is flexible, allowing firms to produce with different technologies, and the method can easily handle multiple outputs and a large number of inputs. The main disadvantages are the requirements on data quality and the assumptions on firm behavior and market structure. It is impossible to account for measurement errors or to deal with outliers, except for some ad hoc trimming of the data. Factor prices information and returns to scale have to be estimated or available independently.

---

[15] The unobservability of capital prices forces me to obtain an outside estimate for returns to scale ($\hat{\epsilon}$) to implement the Törnqvist index number. I estimate returns to scale using least squares and use the came input shares under variable as under constant returns to scale. The productivity growth calculation becomes: $TFPG = \dot{q} - (s^L\hat{\epsilon})\dot{l} - (1 - s^L\hat{\epsilon})\dot{k}$.

[16] The Törnqvist index is just one possibility and different technologies require a different index number. It is the most popular one because it conveniently rationalizes Solow's original TFP formula.

2.3 Parametric methods

The third approach assumes that the input tradeoff and returns to scale are the same for all observations, as in Figure 1. All firm heterogeneity is concentrated in the productivity term.[17] On the plus side, the explicit stochastic framework is likely to make estimates less susceptible to measurement errors. I follow most of the literature by using a simple Cobb-Douglas production function,

$$q_{it} = \alpha_0 + \alpha_l l_{it} + \alpha_k k_{it} + \omega_{it} + \epsilon_{it}, \tag{7}$$

in logarithms. Productivity comparisons are straightforward as the input aggregator in (1) is now assumed constant over time and across firms. Substituting (7) in (1) yields the following productivity comparison

$$\log \frac{A_{it}}{A_{j\tau}} \quad = \quad \log \frac{Q_{it}}{Q_{j\tau}} - \alpha^l \log \frac{L_{it}}{L_{j\tau}} - \alpha^k \log \frac{K_{it}}{K_{j\tau}} - (\epsilon_{it} - \epsilon_{j\tau}), \tag{8}$$

which satisfies $E\Big[\log \frac{A_{it}}{A_{j\tau}}\Big] = \omega_{it} - \omega_{j\tau}$. For some methods it will be possible to calculate the last term in (8) explicitly and subtract the errors from the deterministic component. In other cases, the last term is simply dropped because $E(\epsilon_{it} - \epsilon_{j\tau}) = 0$, in which case the difference in random noise $(\hat{\epsilon}_{it} - \hat{\epsilon}_{j\tau})$ ends up in the productivity term. Equation (8) describes an output-based productivity comparison. It measures the change in output necessary to put one firm on the production frontier of another.[18]

Consistent estimation of the input parameters faces an endogeneity problem, first discussed by Marschak and Andrews (1944). Firms choose inputs, knowing their own level of productivity, which is unobservable to the econometrician. A least squares regression of output on inputs will give biased estimates of the production function coefficients. Three different types of identifying assumptions are implemented to overcome this problem. The most straightforward solution is to use instrumental variables that are uncorrelated with pro-

[17] While it is possible to estimate production functions with random coefficients, allowing for some technology differences between firms, this approach has not been fruitful, see Mairesse and Griliches (1990) for a discussion.

[18] For homogeneous production functions of degree ϵ, it is straightforward to show that the input (A^I) and output (A^O) based measures satisfy the following identity: $\frac{\log A^O}{\log A^I} = \epsilon$.

ductivity. I rely on the method proposed by Blundell and Bond (1998) to generate moment conditions using lagged variables. The stochastic frontier literature makes explicit distributional assumptions about the unobserved productivity factor and estimates the primitives of the distribution. In a more recent approach, Olley and Pakes (1996) obtain an expression for unobserved productivity by inverting the investment function nonparametrically and substitute the expression in the production function. I discuss each of the three approaches in turn.

2.3.1 Instrumental variables estimation (GMM)

A general approach to estimate error component models was developed in Blundell and Bond (1998) and applied to production functions in Blundell and Bond (2000). The model they estimate takes the form

$$\begin{aligned} q_{it} &= \alpha_t + \alpha_l l_{it} + \alpha_k k_{it} + (\omega_i + \omega_{it} + \epsilon_{it}) \\ \omega_{it} &= \rho\omega_{it-1} + \eta_{it} \qquad\qquad |\rho| < 1 \\ & \epsilon_{it},\ \eta_{it} \sim i.i.d. \end{aligned}$$

The production function contains three error components, a firm fixed-effect ω_i, an autoregressive component ω_{it}, with η_{it} an idiosyncratic productivity shock, and measurement error ϵ_{it}. The equation includes year-specific intercepts. In its dynamic representation the model becomes

$$\begin{aligned} q_{it} &= \alpha_l l_{it} + \rho\alpha_l l_{it-1} + \alpha_k k_{it} + \rho\alpha_k k_{it-1} + \rho q_{it-1} \\ &\quad + \underbrace{(\alpha_t - \rho\alpha_{t-1})}_{\alpha_t^*} + \underbrace{\omega_i(1-\rho)}_{\omega_i^*} + \underbrace{(\eta_{it} + \epsilon_{it} - \rho\epsilon_{it-1})}_{\varepsilon_{it}}. \end{aligned} \tag{9}$$

All variables on the first line are observable; firm and year dummies will take care of the first two terms on the second line. There is still a need for moment conditions to provide instruments because the inputs and lagged output will be correlated with the composite error ε_{it}.

Standard assumptions on the initial conditions,

$$\begin{aligned} E[X_{i1}\eta_{it}] &= 0 \qquad \text{for } X = l,k,q \qquad t = 2,...,T \\ E[X_{i1}\epsilon_{it}] &= 0 \qquad \text{for } X = l,k,q \qquad t = 2,...,T \end{aligned}$$

yield three times $T-3$ moment conditions

$$E[X_{it-s}\Delta\varepsilon_{it}] = 0 \qquad \text{for } X = l,k,q. \qquad s = 3,...,T$$

They allow estimation of (9) in first-differenced form using three times lagged inputs and output as instruments. Blundell and Bond (2000) illustrate theoretically and with a practical application that these instruments can be weak. Estimating a production function with firm-dummies or in first differences often leads to very low coefficient estimates, see Mairesse and Griliches (1998).

If one is willing to make the additional assumptions that

$$E[\Delta X_{it}\omega_i^*] = 0 \qquad \text{for } X = l,k \qquad t = 2,...,T$$

and that the initial conditions satisfy

$$E[\Delta q_{i2}\omega_i^*] = 0,$$

one can derive two additional moment conditions

$$E[\Delta X_{it-2}(\omega_i^* + \varepsilon_{it})] = 0 \qquad \text{for } X = l,k. \tag{10}$$

Twice lagged first differences of inputs are valid instruments for the production function (9) in levels. Further lagged differences can be shown to be redundant once the moment conditions in (10) have been exploited. Blundell and Bond (1998) show that joint stationarity of the inputs and output, conditional on common year dummies, is sufficient, but not necessary for (10) to hold.

The GMM-SYS estimator I adopt estimates both versions of the production function, in first differences and levels, as a system with the appropriate set of instruments for each equation. Productivity is calculated by substituting the estimated coefficients in (8), ignoring the error terms.

Advantages of this method are the flexibility in generating instruments and the possibility of testing for overidentification. It allows for an autoregressive component to productivity, in addition to a fixed-effect and an idiosyncratic component. The major disadvantage is the need for a longer panel. At least five years of data are needed to generate overidentifying moment conditions. It is also uncertain, as of yet, how well the instruments work in practice.

2.3.2 Stochastic frontier estimation (SF)

The stochastic frontier literature uses assumptions on the distribution of the unobserved productivity component to separate productivity from the deterministic part of the production function and the random error. The productivity term is modeled as a stochastic variable with negative support. The method is credited to Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977). Stevenson (1980) introduced a truncated normal distribution for ω_{it} that is more flexible on the location of the mode of the distribution. Estimation is usually with maximum likelihood.

In the production function (7), the term ω_{it} is weakly negative and is interpreted as the inefficiency of firm i at time t. The production plan of firm i is said to lie below the best practice production frontier. Alternatively, one can say that firm i produces according to its own production function which is shifted down by ω_{it} relative to best practice.

The original stochastic frontier models were developed to assess productivity in a cross section of firms.[19] The model was subsequently generalized for panel data in a number of ways; I implement two.

Battese and Coelli (1992) provide the most straightforward, but also the most restric-

[19] The same holds for DEA, which is also called deterministic frontier analysis.

tive generalization, modeling the inefficiency term as

$$\begin{aligned} \omega_{it} &= -e^{-\eta(t-T)}\,\omega_i \\ &\text{with } \omega_i \sim N^{+}(\gamma,\sigma^2). \end{aligned} \tag{11}$$

A firm fixed-effect, ω_i, is drawn from a truncated normal distribution and is multiplied by a factor that increases (if η is positive) or decreases (if η is negative) over time. The ranking of firms is unchanged over time and the inefficiency evolves identically and deterministically for all firms. For comparison with the other methods, I calculate productivity according to equation (8), ignoring the last term. The best estimate of $\log A_{it}^{SF1} = E(\omega_{it}|\hat{\omega}_{it}+\hat{\epsilon}_{it})$ is $\hat{\omega}_{it}+\hat{\epsilon}_{it}$ if ω_{it} is independent of ϵ_{it}.[20]

If one observes firms only once, making strong assumptions is the only possibility to separate the productivity component from the random error. Panel data contains more information on each firm and allows identification under weaker assumptions. Schmidt and Sickles (1984) simply use the standard fixed-effects panel data estimator to estimate a constant firm-level productivity term. The problematic correlation between inputs and unobserved productivity has been ruled out by assumption. Cornwell, Schmidt, and Sickles (1990) generalize the method by estimating a time-varying component that is still firm-specific. They adopt a quadratic specification and estimate three coefficients per firm:

$$\omega_{it} = \alpha_{i0} + \alpha_{i1}t + \alpha_{i2}t^2. \tag{12}$$

Firm-level productivity evolves deterministically over time, but the growth rate varies over time and between firms. The productivity level and growth rate can be calculated as

$$\begin{aligned} \log A_{it}^{SF2} - \overline{\log A}_t^{SF2} &= (\hat{\alpha}_{i0} - \overline{\hat{\alpha}_0}) + (\hat{\alpha}_{i1} - \overline{\hat{\alpha}_1})t + (\hat{\alpha}_{i2} - \overline{\hat{\alpha}_2})t^2, \\ \log A_{it}^{SF2} - \log A_{it-1}^{SF2} &= (\hat{\alpha}_{i1} - \hat{\alpha}_{i2}) + 2\hat{\alpha}_{i2}t, \end{aligned} \tag{13}$$

where the overlined variables denote the average over all firms active in year t. Note that

[20] In the stochastic frontier literature, it is customary to estimate technical (in)efficiency as $E(e^{\omega_{it}}|\omega_{it}+\epsilon_{it})$, which is complicated by the nonlinear transformation.

these formulas purge the random errors from the productivity terms.

An advantage of the stochastic frontiers is their relative simplicity to implement. The deterministic part of the production function can be generalized easily, to allow more sophisticated specifications, e.g. incorporating biased technological change. With a short panel, the results for the second estimator are not very robust and that estimator always uses a lot of degrees of freedom. One might also be uncomfortable with the identification coming solely from functional form assumptions, which are especially restrictive in the first specification.

2.3.3 Semi-parametric estimation

The last method was developed by Olley and Pakes (1996) to estimate the productivity effects of restructuring in the U.S. telecommunications equipment industry. They argue that an additional sample selection problem exists if exit is correlated with inputs. If firms exit when productivity drops below a threshold and the exit threshold is decreasing in capital, sample selection will bias the least squares estimate of the capital coefficient downwards.[21]

Olley and Pakes propose a three step estimator to remedy both the selection and endogeneity problem. Under some weak conditions, the investment equation is a monotonically increasing function of productivity and the other state variables—capital and age. The relationship can be inverted, expressing productivity as an unknown function of investment and capital (I ignore age). Substituting that expression in the production function (7) gives the estimating equation for the first step:

$$q_{it} = \alpha_0 + \alpha_l l_{it} + \phi_t(i_{it}, k_{it}) + \epsilon^1_{it}.$$

The unknown function $\phi_t(.)$ is approximated nonparametrically by a fourth order polynomial or a kernel density. In the first step, $\hat{\alpha}_l$ is estimated and $\hat{\phi}_{it}$, which is needed later, is

[21] The general idea of this approach is to use another decision by the firm to provide separate information on the unobserved productivity term. In Olley and Pakes (1996), investment is a monotonic function of productivity. An alternative approach, by Levinsohn and Petrin (2003), inverts the material input equation. An advantage is that firms with zero investment do not have to be dropped from the sample. The estimator in Van Biesebroeck (2003b) inverts another first order condition for automobile assembly plants: the decision how many workers to employ on each shift.

calculated.[22]

The second step deals with the exit decision. Exit is conditional on the realization of productivity and the exit threshold for productivity. Both are different, unknown functions of investment and capital, approximated with a fourth order polynomial on the right-hand side of a probit regression for exit. In the second step, the continuation probability $\hat{P}_{it}$, i.e. the probability a firm remains in the sample, which is needed in the last step, is predicted.

Finally, in the third step, the capital coefficient is estimated. Details on the estimation are in Olley and Pakes (1996), but the intuition is straightforward. From the production function (7), one can write the conditional expectation of $q_{it} - \alpha_l l_{it}$ as $\alpha_0 + \alpha_k k_{it}$ plus the conditional expectation of productivity in period t. If productivity evolves according to a stochastic Markov process, it is a function of its value in the previous period and the exit threshold. Similar to the previous stages, current productivity is approximated nonparametrically using these two variables. The lagged value of productivity can be calculated from the results in the first step as $\hat{\phi}_{it-1} - \alpha_k k_{it-1}$. An expression for the exit threshold can be obtained from the second step, because the continuation probability is a monotonically increasing function of the exit threshold, again an invertible relationship. The estimation equation for the third step is given by

$$q_{it} - \hat{\alpha}_l l_{it} = \alpha_k k_{it} + \psi_t(\hat{\phi}_{it-1} - \alpha_k k_{it-1}, \hat{P}_{it-1}) + \epsilon^2_{it}.$$

Only the capital coefficient is left to estimate at this stage.

Once the coefficients in the production function are estimated, productivity is calculated from (8) ignoring the last term, as in Olley and Pakes (1996). These results are indexed by $OP1$. It is also possible to calculate a direct estimate of $\hat{\omega}_{it}$, purged from random noise $\hat{\epsilon}_{it}$, as

$$\log A^{OP2}_{it} - \log A^{OP2}_{j\tau} = (\hat{\phi}_{it} - \hat{\alpha}_k k_{it}) - (\hat{\phi}_{j\tau} - \hat{\alpha}_k k_{j\tau}). \tag{14}$$

This measure can only be calculated for firms with positive investment, i.e. the firms included

[22] When constant returns to scale is enforced, only the first stage is estimated, and $\hat{\alpha}_k = 1 - \hat{\alpha}_l$.

in the estimation procedure.

The main advantage of this approach is the flexible characterization of productivity. The only assumptions are the Markov process for the evolution of productivity and that the nonparametric approximations are adequate. The investment function to be inverted is likely to be a complicated mapping from states to actions since it has to hold for all firms regardless of their size or competitive position. It is unsure how good an approximation the fourth order polynomials or kernel densities provide in a small sample.

3 Direct Comparison of Methodologies

3.1 A panel of Colombian manufacturing plants

Even though the calculations for each method differ substantially, they all intend to compare output differences, while controlling for input differences, as in equation 1. It is therefore uncertain how much the estimates will differ in practice. I evaluate the different methodologies using a sample of manufacturing plants from Colombia. In this section, the productivity level and growth estimates are compared directly. In the next section, estimates from all methodologies are used to revisit three debates that received a lot of attention in the development economics literature.

The data comes from the annual census of manufacturing, which covers all active establishments, between 1977 and 1991. Only establishments that are classified in the ISIC (Revision 2) 322 industry—Clothing and Apparel—at some point during the sample are included.[23] The sample is further limited by only including plants that operate for at least three years, as many estimation methods need at least three observations per plant. This results in an unbalanced panel of 14348 observations from 1957 plants with nonmissing information on output, labor and capital input, wages, and investment. 8% of observations employ 10 or less employees, 13% employ 100 or more. More information about the data and variable

[23]This industry was chosen as plants are expected to be relatively homogeneous in technology and because this sector has a large foreign exposure, which will be important in the debates. At the end of the sample period, this industry accounts for 10% of manufacturing employment, 3% of value added, and 8% of exports.

construction can be found in Roberts (1996). Table 1 contains some summary statistics.

[Table 1]

The output concept used is value added, defined as sales minus indirect costs and material input. Labor input is total employment and capital input is the reported book value of the plant and equipment. Value added is deflated with the same sectoral output deflator as used in Roberts (1996). For capital, the capital goods deflator from the IMF Financial Tables is used.

3.2 Estimation results

Productivity levels and growth rates are estimated using each methodology, under constant and variable returns to scale. The table below summarizes the superscripts used for the different estimation methods. The coefficient estimates for the production function parameters

index	method	(equation)
OLS	Least squares estimation of VA on L and K (benchmark)	(8)
IN	Törnqvist Index with correction for returns to scale	(6)
DEA	Data Envelopment Analysis, pooling all years	(3)
GMM	GMM-SYS estimation of equation (9)	(8)
SF1	Stochastic frontier, productivity as in equation (11)	(8)
SF2	Stochastic frontier, productivity as in equation (12)	(13)
OP1	Semiparametric, as in Olley and Pakes (1996)	(8)
OP2	Estimation as in OP1, productivity calculated differently	(14)

and comparable statistics for the Törnqvist index and DEA analysis are in Table 2.

[Table 2]

If constant returns to scale are enforced the parameter estimates are quite similar for the different parametric methods. Relative to the OLS estimates, accounting for the simultaneity bias lowers the labor coefficient and increases the capital coefficient, with the largest change for the GMM estimator. All estimates are very precise and standard errors are omitted.[24] The index number and DEA calculations allow for heterogeneity in technology for different observations. The average labor coefficients are substantially lower than with the parametric methods. This might suggest that the corrections for the endogeneity of productivity in the factor input choices that the parametric methods employ are only partially successful.

If returns to scale are estimated freely, the differences between the different methodologies become more pronounced. The OLS estimate of the labor coefficient increases slightly, with no change for capital. This results in small, but significantly higher than unity, returns to scale (RTS). The index number calculations use the RTS estimate from OLS for all observations—as an exogenous estimate is needed.[25] The DEA method finds remarkably similar average RTS, but a relatively lower weight for capital. The weights vary less across observations under VRS. The standard deviation of the labor weight across all observations drops from 0.39 to 0.30.

The coefficient estimates with the parametric methods look altogether more troublesome. The adjustments for the endogeneity of productivity tend to reduced the labor coefficient estimate more than the capital coefficient increases, relative to OLS. Returns to scale are invariable estimated to be decreasing, which is unexpected for manufacturing plants in a developing country. Such result are fairly common when the production function is estimated in first-difference form (as in GMM) or with a full set of plant-dummies (as in SF2), see Mairesse and Griliches (1998). With the other two methods, SF1 and OP, scale economies are estimated closer to unity. One consistent finding is that the capital coefficient is relatively more important than in the least squares or the CRS results, as expected.

[24] More detailed estimation results are available from the author upon request.

[25] Monte Carlo simulation results in Van Biesebroeck (2003c) show that OLS estimates RTS relatively accurately when measurement error or the extent of simultaneity bias varies. The labor coefficient is overestimated, but a corresponding underestimation of the capital coefficient leaves the sum largely unaffected.

3.3 Productivity levels

A first way to compare the productivity estimates is to compare the dispersion they imply. The first two columns in Table 3 contain the interquartile range under constant returns to scale for each method. The median is normalized to zero by year. The ranges are very similar, which is remarkable because the methods rely on very different calculations and assumptions. Dispersion appears to be relatively large in Colombia. On average, only half of the firms have a productivity level between 32% below and 35% above the median firm in the sample. The two nonparametric methods, IN and DEA, allow for the technology to differ between plants and they find intervals that are approximately 25% wider than for the parametric methods. All methods find the distribution of productivity to be slightly skewed to the right. Even the SF1 methodology finds this, while left-skewness is built in, as productivity is the sum of a symmetric normal error and a "productivity" term with a negative half-normal distribution. An underlying economic model with a frontier technology that firms aspire to (SF1) appears to be less plausible than a model where firms are forced to exit when their productivity drops below a threshold (OP), cutting off the lower tail of the distribution.

The one anomalous range is for the OP2 method. It deducts the random measurement error from the OP1 results, to obtain a pure productivity estimate, while OP1 results reflect the sum of productivity and measurement error ($\hat{\omega} + \hat{\epsilon}$). The OP2 estimate should be interpreted as the firm's own estimate of its productivity level that it takes into account when choosing its inputs. Idiosyncratic productivity shocks that are realized only after inputs are chosen will show up in the OP1 estimates but not in OP2. The semiparametric approach suggests that only a small part of the estimates that the other methods come up with can be identified as true underlying productivity, while most of it is random error or idiosyncratic productivity shocks.

The results under VRS generally confirm the previous findings. The methods that estimated RTS to be very low (GMM and SF) and the DEA method see a substantial widening of the interquartile range, with much smaller and mixed changes for the other methods. The right-skewness also becomes more pronounced. Again, the interquartile range for OP2 is much smaller than for the other methods, while the SF2 methodology—the only other method that explicitly purges the regression error from the productivity estimates—produces one of the widest intervals.

[Table 3]

Another way to compare the methodologies is to look at the correlations between the productivity level estimates. Table 4 contains the correlations both under CRS (in the bottom-left) and under VRS (in the top-right). Only the OP2 calculations produce results that differ substantially from the other methods. This could be expected from the previous discussion, as it is really estimating something else. The OP2 method does generate the highest correlation with the SF2 method, the only other one that subtracts measurement error, especially when returns to scale are estimated freely.

The results imposing constant returns to scale produce more highly correlated productivity estimates than the variable returns results. Apart from the OP2 results, the lowest correlation under CRS is 0.79. Even the DEA method, that leaves technology completely unrestricted, produce highly similar estimates. The two nonparametric methods—DEA and IN—produce results most alike each other. The parametric methods that use equation (8)—OLS, GMM, SF1, OP1—are virtually indistinguishable under CRS.

Allowing scale economies or diseconomies generates more disparate measures, but even then only correlation coefficients with OP2 are ever below 0.50. The SF2 results become less highly correlated with the other methods, while the GMM method that estimates RTS equally low still produces results that are very similar to other methods. While the index number and DEA methods are still most alike each other, they still resemble the parametric results remarkably well, even though the range of RTS estimated is quite large. The estimates from the four parametric methods—omitting OLS—all produce correlation coefficients between 0.75 and 0.99.[26]

[Table 4]

[26] Spearman-rank correlations between the different methods are similar, but slightly lower for the stochastic frontier estimates. Calculating the correlations separately by year, yields virtually identical results.

3.4 Productivity growth

To compare the productivity growth estimates, Table 5 lists the unweighted and aggregate input weighted averages for the entire sample. The period from 1977 to 1991 was a very successful one for Colombian clothing plants. The average growth rate, across all methods but OP2, is 6% per firm per year under CRS and 5.6% under VRS. By all standards, these are very high numbers for total factor productivity growth. The two methods that take out measurement error, SF2 and OP2, produce the lowest estimates, indicating that whatever they subtract—measurement error or something else—was trending upwards. The unweighted average growth rates are estimated very similar across methods, again omitting OP2, differing by at most 0.6% under CRS and 1.5% under VRS.

A different pictures emerges when plants are weighted. The issue of correct weights in aggregation is revisited below when aggregate productivity growth is decomposed. Using input weights will generally lower the average, as plants that improve productivity most—those that decreased input use relative to output—end up with low inputs and receive a low weight. The reverse happens with output weights; plants with high output growth have, *ceteris paribus*, high productivity growth and receive a higher weight. An advantage of denominator (input) weights is that the average approximates the growth rate that one obtains from aggregate input and output statistics. For labor productivity, aggregate growth is reproduced exactly. An input aggregate is used as weight in Table 5.[27]

Under CRS, four parametric methods—OLS, GMM, SF1, OP1—see an almost equal drop in average growth rate to 2.7% per year. Under VRS, they differ somewhat more, but still experience a similar reduction in mean growth rate, on average to 3.5%. The two methods that take out measurement error—SF2 and OP2—are less affected by weighting, both under CRS and VRS. Moreover, they produce similar results under CRS and VRS. The results from

[27] An input-aggregate is calculated as follows:

$$\begin{aligned} \theta_{it}^{Z} &= \frac{Z_{it}}{\sum_{j=1}^{N} Z_{jt}} \\ \text{with } Z_{it} &= L_{it}^{\hat{\alpha}_L} K_{it}^{\hat{\alpha}_K} \end{aligned}$$

IN and DEA differ to a greater extent, but that could be expected. These two methods allow different input weights across plants when productivity is calculated, while the aggregate input weight necessarily uses the same weights for all plants. When output weights are used they generate similar results as the other methods (results available upon request). In sum, the different methods paint a very similar picture about the average productivity growth that these plants experienced.

[Table 5]

Finally, the correlations of productivity growth estimates between methods, follows the pattern of the level correlations closely. The OLS, GMM, SF1, and OP1 methods produce virtually identical results under CRS. Only the GMM method produces slightly different results under VRS—recall the very low estimate for RTS—, but even then correlation coefficients are around 0.95. The DEA and IN methods still resemble each other most, but produce results that are very close to the parametric results as well. The differences in input coefficients are swamped by the huge differences in output and input growth rates across firms. The OP2 results are hardly correlated with any other method. The part of productivity known to the firm is more stable than the productivity growth that other methods come up with, see results in Table 5. The changes are likely to be swamped by idiosyncratic productivity shocks. The growth estimates obtained by the SF2 method differ more from other methods than the level estimated did. OP2 estimates are equally similar to the parametric as to the nonparametric results.

[Table 6]

The results in Tables 5 and 6 suggest that the different methods are even more alike when productivity growth rates are calculated than for productivity levels. Especially, the similarity between the nonparametric and parametric results is remarkable.

4 Three Debates

A completely different approach to evaluate how similar the productivity estimates are is to investigate whether the answers to a number of debates are sensitive to the specific measurement methodology used. I revisit three debates using each methodology and summarize the Colombian experience.

4.1 Have exporters higher productivity levels?

Many studies have found that exporting plants have higher productivity, see for example Bernard and Jensen (1995) and Aw, Chen, and Roberts (2001). The direction of causality, on the other hand, is still debated. Some authors argue that the positive correlation is solely the result of self-selection of the most productive producers into the export market, see Clerides, Lach, and Tybout (1998) and Bernard and Jensen (1999). Others argue that there is a role for learning-by-exporting effects. Firms that are exposed to foreign competition will benefit and improve productivity, see for example Kraay (1999) and Van Biesebroeck (2003a).

Equally well established is the notion that exporters are larger than average firms. This finding is robust across continents; evidence for the United States is in Bernard and Jensen (1995), for Colombia in Roberts and Tybout (1997), and for Africa in Van Biesebroeck (2003a). One benefit of exporting, especially in developing countries, is the ability to realize scale economies. Tybout (2000) summarizes the evidence and concludes that scale economies in the manufacturing sector of developing countries ranges from 1.05 to 1.1. It is an open question, what percentage of the productivity advantage that exporters enjoy can be attributed to their size.

I calculate the productivity premium for exporters under different assumptions on returns to scale. If scale economies are present and exporters are larger, the productivity effect associated with export status will be estimated higher under CRS. The export dummy will soak up some of the scale effect because exporters are undeniably larger on average. If the coefficient is unchanged when scale economies are estimated freely, size is immaterial in explaining the export premium. The reduction of the export status coefficient if productivity is calculated allowing VRS can be interpreted as the part of the export premium that is explained by scale effects.

Not surprisingly, the results in Table 7 are for a large part driven by the disparity in returns to scale estimates. The statistics in the first column are coefficient estimates on an export dummy in separate regressions. The dependent variable is log-productivity, calculated under the CRS assumption, and a number of controls are included. All coefficients are estimated to be positive and significantly different from zero. All parametric methods find virtually the same productivity premium for exporters, on average 31%. The OP2 method finds a smaller advantage in absolute value, but attributes to them a productivity premium that would place the median firm above the 75th percentile in the productivity distribution.

The two nonparametric methods—IN and DEA—estimate a substantially lower productivity premium. Recall that these methods allow different input coefficients for different plants. Previous studies have shown that exporters are not only larger, they also produce with a larger capital stock per employee. Given that only few plants export, 9% in this sample, the estimated input coefficients for the parametric methods are more representative of non-exporters' production plans. Properly accounting for the higher capital intensity of exporters explains more than half of the estimated productivity advantage, reducing it to 13%.

All coefficients remain positive and significant when returns to scale are allowed to vary, but the results vary a lot by method. The OLS results are consistent with our prior. The larger size of exporters and the estimated increasing returns to scale combine to reduce the productivity premium estimate by almost one quarter. The index numbers, on the other hand, estimate a larger productivity premium under VRS, even though returns to scale are increasing. The DEA result, which allow firm specific scale economies as well as input weights, finds a very high productivity premium for exporters. Finally, the parametric methods found decreasing scale economies, which lead them to interpret the larger size of exports as even higher productivity advantages. For the GMM and SF1 results, the implausibly low estimates for scale economies lead to implausibly high productivity premiums for exporters.

In the third and fourth column of Table 7, I use lagged instead of contemporaneous export status in an attempt to control for the self-selection of more productive plants into the export market. Unfortunately, the large persistence in export status, potentially due to sunk export costs as argued in Roberts and Tybout (1997), makes it a a weak control at best.[28] With few exceptions, the productivity premium only drops slightly. Rather than addressing the tricky question of causality, I would like to draw attention to the similarity in the changes. Every coefficient estimate decreases by a small amount, 1.1 to 4.3%, which is very similar across methods. The only instances where the productivity premiums increase are for the two methods that subtract the error term—SF2 and OP2. Self-selection of plants into the export market due to idiosyncratic productivity shocks is consistent with a higher productivity premium for current relative to lagged export status for the other methods and not for these two.

Results in the last two columns of Table 7 add the share of sales exported to the equation. It reveals that the productivity premium exporters enjoy is certainly not proportional to the amount of exporting they do. Higher export shares are associated with correspondingly lower productivity premiums. Again, the results are very consistent across methods. It mirrors findings for Colombia in Isgut (2001) and contrasts with the positive estimates previous studies have found for the U.S. and Germany.

In sum, all methods find that exporters are more productive and that the productivity advantage cannot be explained solely by scale economies. The extent to which the premium is due to scale economies, on the other hand, is sensitive to the specific method used to estimate productivity. The results also suggest that assuming all plants produce with the same technology, in particular that they face the same capital-labor tradeoff, is likely to lead to an overestimate of the productivity premium for exporters.

[Table 7]

[28] The correlation between lagged and current export status is 0.76.

4.2 What brings about technological change?

The main trust of the growth model in Solow (1956) was that long term per capita growth can only come from technological change. The source of technological change was left unspecified, exogenous to the model. Capital accumulation can only increase output in the transition to the long run, until it runs out of steam because of diminishing marginal return. In later developed endogenous growth models, e.g. Romer (1986), the additional knowledge is created *endogenously* in the economy by profit maximizing firms. Jones (1995) tests the endogenous growth model using U.S. data, but did not find support for it.

In developing countries, the amount of domestic R&D is negligible. The main engines of growth that researchers have investigated are improvements in physical and human capital and imported knowledge from more advanced economies. While capital accumulation cannot account for long term economic growth, it can be the vehicle for embodied technical change to enter the economy. Computers are often thought to perform this task in developed countries, see Brynjolfsson and Hitt (2003) for recent evidence. In developing countries, imported machinery is often important. Tybout (2000) summarizes some of the evidence. Many factors are found to have a positive correlation with the productivity level. Foreign owned plants and exporters are more productive; firms that employ many engineers, scientist, and technical workers are more productive; high rates of investment in fixed equipment improves productivity, see De Long and Summers (1991) for an influential investigation at the aggregate level.

While in the long run the effects on productivity growth are more important than level effects, they have also proven to be more elusive. The correlation of many variables with the level of productivity can be misleading as it might simply capture differences in input quality. Productivity growth comparisons are less susceptible to this criticism. In Table 8, the growth effects of international exposure is investigated. In Table 9, the same exercise is performed for different measures of human and physical capital investment. In both tables, the average growth rate of a plant over the entire period it is active is regressed on a dummy variable. Each coefficient is estimated in a separate regression.

In Table 8, only the export dummy consistently has some explanatory power. Plants that export improve their productivity at a faster rate, regardless of the measurement methodology used. On average they record 2% higher productivity growth, which is significantly different from zero for the parametric methods. The effect of the export dummy is often interpreted as a spillover effect. Competing with foreign firms for business or having to satisfy demanding foreign clients exposes firms to best practice methods and forces firms to improve their own production technology. How this happens is generally not specified. It was discussed before that firms with a high productivity level might self-selected into the export market. It cannot be ruled out that firms with high productivity growth self-select similarly.

Productivity improvements does not seem to happen through more advanced inputs, foreign workers, or direct payments for new technology. These other characteristics are not informative, producing ambiguous and mostly insignificant effects. Paying royalties, importing inputs, or employing foreign workers have a negative, if any, impact on productivity growth. This is consistent with the model in Acemoglu and Ziliboth (2001). They show that many technologies used in developing countries are developed in OECD countries and are inappropriate for the local mix of skills. Having access to new technologies will not suffice to improve productivity in developing countries.

[Table 8]

The results in Table 9 look more promising. The first columns indicate that investments in new production or office equipment (physical capital) tends to increase productivity growth. Each dummy is defined to take on the value of one for one third of the plants: in the first column, for the third of plants with the highest average investment as a percentage of capital stock (i.e. plants whose yearly investments average at least 17% of their capital stock); in the second column, for plants with the highest one-year investment spike (higher than 60%); in the third column, for plants that made investments most frequently (at least 70% of the years that they were active). The results indicate that it is more important to invest frequently than investing a lot. For most methods, frequent investments are strongly correlated with high productivity growth. Plants that make frequent investments are rewarded with 2.3% higher productivity growth, on average. The results for high average or peak investments tend to confirm this finding, but with some exceptions and the effects are uniformly smaller.

The last three columns in Table 9 hint at a human capital effect. The "managers" column contains the coefficient of a regression of average productivity growth on a dummy that takes on the value of one if the plant is in the top one-third of plants in terms of the share of managers in total employment. All but one coefficients are positive and five are significantly different from zero. The results are also fairly homogeneous. The parametric methods point to a 2% productivity growth advantage, while the nonparametric methods do not find any effect. A glance at the last two columns reveals that the results completely disappear or even turn negative for the other two worker categories that would be associated with high levels of human capital. The other, omitted, worker categories are owners not paid a fixed wage, unskilled workers, and apprentices. It is noteworthy that the average salary difference between skilled and unskilled workers are very small. The skilled denomination might not proxy human capital differences very well.

While few variables are consistently associated with above average productivity growth, the different methods are in in close agreement. Export status, frequent investments, and a high fraction of managers in the workforce are associated with high growth. While the effects are rarely significant, paying royalties, employing foreign, high skilled, or technical workers, is associated with lower growth.

[Table 9]

4.3 What drives aggregate productivity growth?

A final debate concerns the decomposition of aggregate productivity growth into the contribution of plant-level changes in productivity and the reallocation of input or output shares between plants. Obviously, the aggregate productivity level increases if individual plants become more productivity. At the same time, it is possible for aggregate productivity to increase without any plant-level productivity growth, if inputs are moved from plants with a below average productivity level to plants with above average productivity. The question is then: which of the two effects is most important in practice?

To my knowledge, the first decomposition using the universe of plants was performed by Baily, Hulten, and Campbell (1992) for the United States. Haltiwanger (1997) introduced an improved formula for unbalanced panels with an extra covariance term:

$$\begin{aligned}\Delta TFP_t \;=\; TFP_t - TFP_{t-1} \;=\; & \sum_{n=j,k}^{\text{stay, enter}} \theta_{nt}\, TFP_{nt} - \sum_{m=j,l}^{\text{stay, exit}} \theta_{mt-1}\, TFP_{mt-1} \\ =\; & \sum_{j}^{\text{stay}} \left[\theta_{jt-1}\, \Delta TFP_{jt} + \Delta\theta_{jt}\,(TFP_{jt-1} - TFP_{t-1}) + \Delta\theta_{jt}\, \Delta TFP_{jt} \right] \\ & + \sum_{k}^{\text{enter}} \theta_{kt}\,(TFP_{kt} - TFP_{t-1}) - \sum_{l}^{\text{exit}} \theta_{lt-1}\,(TFP_{lt-1} - TFP_{t-1}). \qquad (15)\end{aligned}$$

TFP_{it} is the logarithm of productivity calculated using the different methods introduced before, without the normalization. Aggregate productivity growth is defined as $TFP_t - TFP_{t-1}$, the difference between the aggregate productivity levels, which are defined as $\sum_i \theta_{i\tau} TFP_{i\tau}$. Plants that stayed in the sample from $t-1$ to t are indexed by j. Their contribution is split into three terms. The first term measures the total effect of plant-level productivity changes, weighted by the initial share. The second term captures the reallocation effect; it sums changes in shares using a plant's productivity relative to the average productivity level in the initial period as weight. If a plant with above average productivity becomes larger, the contribution will be positive. The third term captures the covariance between the plant-level growth and reallocation. In the original decomposition by Baily, Hulten, and Campbell (1992) the covariance term was lumped with the reallocation term, even though it captures both within plant and between plant effects. Plants that enter in t are indexed by l and contribute positively if they have higher productivity than the aggregate in the previous period. Plants that are last observed in the sample in $t-1$ are indexed by k and their contribution is subtracted.

Most studies have used output-weights. Bartelsman and Dhrymes (1998) correctly point out that it is more intuitive to use an input aggregate as weight.[29] For labor productivity, weighing individual measures by labor shares exactly reproduces aggregate productivity.[30] For TFP, it is still impossible to reproduce the aggregate from the individual measures.[31]

Baily, Hulten, and Campbell (1992) concluded that the bulk of growth in aggregate TFP was accounted for by reallocation of output shares. More productive plants gradually capture a larger market share. Alternatively, Bartelsman and Dhrymes (1998) plot aggregate TFP growth against the simple average of plant-level TFP growth for the same sample. They find that the unweighted average is almost constant over time, while the weighted average increases substantially and also conclude that reallocation effects dominate. The modified decomposition by Haltiwanger (1997) revealed that it was the covariance effect that was responsible for the reallocation effect. The comparison with the Colombian results from Tybout and Liu (1996) should be done cautiously as they used the old decomposition, input-weights, and looked at year by year changes. The positive contribution of the combined term might mask a negative between effect and a positive covariance.

New results for the Colombian textile industry, using aggregate input weights, are in Table 10. The first column contains the cumulative change in aggregate productivity level over the 1977-1987 period, calculated using the productivity estimates from each method, enforcing constant returns to scale. The next five columns decompose the aggregate into the five terms of equation (15). Comparable results using output weights are in Table 11.

The index numbers results were omitted, because the heterogeneity of technology makes

---

[29] $\theta_{it}^{Z} = \dfrac{Z_{it}}{\sum_j Z_{jt}}$, with $Z_{it} = L_{it}^{\hat{\alpha}^L} \; K_{it}^{\hat{\alpha}^K}$.

[30] $LP_t = \dfrac{\sum_i Q_{it}}{\sum_i L_{it}} = \sum_i \left(\dfrac{L_{it}}{\sum_j L_{jt}} \right) \dfrac{Q_{it}}{L_{it}} = \sum_i \theta_{it}^{L} LP_{it}$

[31] Aggregate productivity is calculated as $TFP_t = \dfrac{\sum_i Q_{it}}{\left(\sum_i L_{it}\right)^{\hat{\alpha}^L} \left(\sum_i K_{it}\right)^{\hat{\alpha}^K}}$, while the sum of input-weighted plant-level productivity produces $TFP_t^* = \dfrac{\sum_i Q_{it}}{\sum_i \left(L_{it}^{\hat{\alpha}^L} K_{it}^{\hat{\alpha}^K} \right)}$. The two formulas differ by the order of summation and geometric weighting.

	Country	Period	Total	Within	Between	Covariance	Net entry
(1)	U.S.	1977-87	10.7	5.8	-1.1	4.0	2.0
		1977-82	2.4	-0.3	-1.3	3.5	0.4
		1982-87	8.3	4.8	-1.4	3.9	1.0
(2)	Colombia	1977-86*	1.0	1.0	1.0		-1.0
		1977-82*	-4.4	-2.5	0.3		-2.3
		1982-86*	5.4	3.5	0.7		1.3

(1) Haltiwanger (1997); all industries; output weights

(2) Tybout and Liu (1996); all industries; aggregate input weights.

* sum of year-by-year changes.

it impossible to fit them in the decomposition formula. To make the different terms sum up, productivity growth at the plant level should be redefined, because it is not simply the log difference between the productivity level estimates. Even that fix would be inappropriate because input weights differ by firm and are year specific. The resulting productivity level or growth comparisons would not be not invariant to the unit of measurement, e.g. measuring the capital stock in pesos or thousands of pesos would produce different results.

In both tables, the results are remarkably similar for all but one method. The OP2 method produces different results, but we know from before that it measures a different concept. It is no surprise that it generates different findings when we disentangle within and between firm effects. In the bottom rows, the same decompositions are performed for labor productivity. Here, input (labor) weights are clearly preferable. The results mirror the findings for total factor productivity very closely.

Using input weights, in Table 10, the majority of the cumulative change in aggregate productivity is caused by changes at the plant level. All methods produce very similar numbers, around 29% or between two thirds and three quarters of the total changes comes from within plant changes. The second most important effect, by far, is the entry of more productivity plants into the economy. Note that they are not necessarily more productive at the time of entry. They enter some time between 1977 and 1987. Conditional on surviving till the end of the period, they are more productive than the average plant was in 1977. All methods estimate the importance of entry between 40% and 50% of the total effect. These findings confirm the importance of net entry for the U.S. when a longer time horizon is considered. They contrast with U.S. result in Haltiwanger (1997) on the plant level effects. However, a recent study on the U.S. textile industry finds a similar importance of within plant changes, see Levinsohn and Petropoulos (2001).[32]

The other three terms have uniformly negative and much smaller contributions. Inputs tend to flow from more to less productive plants and this effect is most pronounced for the labor productivity growth decomposition. Plants that tended to employ a lot of workers in 1977 for the output they produced and that survived till the end employ even more workers in 1987. Plants that exited between 1977 and 1987 were on average more productive than the average plant in 1977 and their exit lowers productivity growth. This is in contrast with findings for other, more developed, countries. The market in Colombia does not seem to do a good job wielding out less efficient producers. Because the exit effect is smaller than the entry effect, net entry contributes positively to aggregate growth. Finally, the covariance term is relatively unimportant and negative for all methods except OP2. Plants that improve productivity tend to reduce their input share and plants that deteriorate tend to increase their share in input use. It indicates that outputs and inputs move in opposite directions or that input changes dominate output changes. A positive contribution is only possible if higher input use is accompanied by even more rapid output growth or if a declining input share leads to negative productivity growth because output declines more than proportionally. Both phenomena are not very common. Comparison with the U.S. has to wait, because it really matters which weights are used for the covariance term.

[Table 10]

Using output weights, in Table 11, the results only differ substantially for the covariance term. The results from each method are equally close. Only the OP2 method produces anomalous results, as before, and won't be discussed further.

Aggregate growth is estimated slightly lower for every method. The contribution of plant-level changes and net entry are still the two most important effects. The within plant

[32] Because the different terms fail to sum to aggregate productivity growth, I did not include these results in the table above.

component is slightly less important, on average accounting for 62% instead of 73% of the total growth, while the entry effect is more pronounced, contributing 54% versus 45% of the total.

The negative between component indicates that surviving low productivity plants grab a larger share of the market in 1987 than in 1977. It can be indicative of markets failing to allocate resources efficiently or it can capture plants converging to the industry average. In the period from 1977 to 1987, the economy was liberalized and deregulated substantially. The production index for the textile sector increased by only 7.5%, while the average was 37% and a third of all three digit industries saw output jump by more than 50%. The exit term still contributes negatively to aggregate growth. Plants that exit had above average productivity level in 1977. Note that a firm exits the sample only if it does not show up anymore in the census, not if it changes industry. Net entry is still positive, but uniformly smaller than using input weights.

Finally, the one term that changed most is the covariance term. The correlation between productivity growth and output growth is positive. Plants that improve productivity also increase their output share; or, alternatively, plants that deteriorate productivity lose market share. Market share does not increase for plants that were above average productive last period, but it does increase for plants that increased their productivity over the period. This is consistent with all inputs remaining put, and output alone shifting. As were the results in Table 10 consistent with solely inputs moving between plants without any output changes. In reality, both output and input shares change and each decomposition stresses one of the two effects. The contribution of the covariance term is estimated to be just over 10% with each methods.

[Table 11]

In both tables, the different methods produce very similar results. The choice of weights turns out to be more important for the results than the choice of estimation method for productivity.

5 Lessons

To conclude, in response to the question "Does it matter which method we use to estimate productivity," I answer with a qualified no. Only the method that calculates a distinct underlying concept—OP2—produces different productivity level and, especially, growth estimates. The only other method that explicitly takes out random measurement error—SF2—also produces somewhat different estimates (except for productivity levels enforcing constant returns to scale). Correlations, interquartile ranges, and averages are very close for all other methods. Even the deterministic DEA and index number approaches generate similar results.

I reach the same conclusion when revisiting the three productivity debates. The choice of estimation method for productivity is largely immaterial to the conclusions reached.

- Exporters have higher productivity using each method. The extent to which this is due to scale economies, on the other hand, differs tremendously by method, as the returns to scale are estimated very low for some parametric methods.

- The association between productivity growth and foreign exposure is weak at best. It is positive for exporters, possibly due to self-selection, but all other foreign exposure dummies have no or a negative impact on productivity growth, using each method. The association between productivity growth and human or physical capital investments is slightly stronger. Most methods find a positive and significant effect from frequent capital investments and the presence of many managers.

- Using input or output weights, all methods except OP2 find large and positive contributions of plant-level changes and net entry on aggregate productivity growth. The between firm component is negative, while the sign and interpretation of the covariance term depends on the weights. Again, all methods except OP2 result in extremely similar decomposition results.

References

Acemoglu, D. and F. Ziliboth (2001, May). Productivity Differences. *Quarterly Journal of Economics 116*(2), 563–606.

Aigner, D., C. K. Lovell, and P. Schmidt (1977). Formulation and Estimation of Stochastic Frontier Production function Models. *Journal of Econometrics 6*, 21–37.

Aw, B. Y., X. Chen, and M. J. Roberts (2001). Firm-level Evidence on Productivity Differentials and Turnover in Taiwanese Manufacturing. *Journal of Development Economics 66*, 51–86.

Baily, M. N., C. Hulten, and D. Campbell (1992). Productivity Dynamics in Manufacturing Plants. *Brookings Papers: Microeconomics 4*(1), 187–267.

Bartelsman, E. J. and P. J. Dhrymes (1998). Productivity Dynamics: U.S. Manufacturing Plants, 1972-1986. *Journal of Productivity Analysis 9*, 5–34.

Battese, G. E. and T. J. Coelli (1992). Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India. *Journal of Productivity Analysis 3*, 153–69.

Bernard, A. B. and J. B. Jensen (1995). Exporters, Jobs, and Wages in U.S. Manufacturing, 1976-1987. *Brookings Papers on Economic Activity, Microeconomics*, 67–119. Washington DC.

Bernard, A. B. and J. B. Jensen (1999). Exceptional Exporter Performance: Cause, Effect, or Both? *Journal of International Economics 47*(1), 1–26.

Berndt, E. R. and M. A. Fuss (1986, Oct./Nov.). Productivity Measurement with Adjsutments for Variations in Capacity Utilization and Other Forms of Temporary Equilibrium. *Journal of Econometrics 33*(1/2), 7–29.

Blundell, R. W. and S. R. Bond (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics 87*, 115–43.

Blundell, R. W. and S. R. Bond (2000). GMM Estimation with Persistent Panel Data: An Application to Production Functions. *Econometric Reviews 19*(3), 321–340.

Brynjolfsson, E. and L. M. Hitt (2003, forthcoming). Computing Productivity: Firm-Level Evidence. *Review of Economics and Statistics*.

Caves, D. W., L. R. Christensen, and E. W. Diewert (1982a). The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity. *Econometrica 50*(6), 1393–1414.

Caves, D. W., L. R. Christensen, and E. W. Diewert (1982b). Multilateral Comparisons of Output, Input, and Productivity using Superlative Index Numbers. *Economic Journal 92*, 73–86.

Charnes, A., W. W. Cooper, and E. Rhodes (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research 2*, 429–444.

Clerides, S. K., S. Lach, and J. R. Tybout (1998). Is Learning By Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco. *Quarterly Journal of Economics 113*(3), 903–947.

Coelli, T., D. P. Rao, and G. E. Battese (1997). *An Introduction to Efficiency and Productivity Analysis*. Norwell, MA: Kluwer.

Cornwell, C., P. Schmidt, and R. C. Sickles (1990). Productivity Frontiers with Cross-sectional and Time Series Variation in Efficiency Levels. *Journal of Econometrics 46*, 185–200.

De Long, B. and L. Summers (1991). Equipment Investment and Economic Growth. *Quarterly Journal of Economics 106*, 445–502.

Denison, E. F. (1972). Some Major Issues in Productivity Analysis: an Examination of the Estimates by Jorgenson and Griliches. *Survey of Current Business 49*(5, Part II), 1–27.

Denny, M., M. Fuss, and L. Waverman (1981). The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with an Application to Canadian Telecommunications. In T. Cowing and R. Stevenson (Eds.), *Productivity Measurement in Regulated Industries*, pp. 179–218. New York: Academic Press.

Diewert, W. E. (1976). Exact and Superlative Index Numbers. *Journal of Econometrics 4*, 115–145.

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society, Series A. 120*, 253–290.

Griliches, Z. and V. Ringstad (1971). *Economies of Scale and the Form of the Production Function*. Amsterdam: North Holland.

Haltiwanger, J. C. (1997). Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence. *Federal Reserve Bank St. Louis Review 79*(3), 55–77.

Isgut, A. E. (2001, June). What's Different about Exporters? Evidence from Colombian Manufacturing. *Journal of Development Studies 37*(5), 57–82.

Jones, C. I. (1995, May). Time Series Tests of Endogenous Growth Models. *Quarterly Journal of Economics 110*, 495–525.

Jorgenson, D. W. and Z. Griliches (1967, July). The Explanation of Productivity Change. *Review of Economic Studies 34*, 349–83.

Klette, T. J. and Z. Griliches (1996). The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous. *Journal of Applied Econometrics 11*(4), 343–361.

Kraay, A. (1999). Exportations et Performances Economiques: Etude d'un Panel d'Entreprises Chinoises. *Revue d'Economie Du Developpement 0*(1-2), 183–207.

Levinsohn, J. and A. Petrin (2003, April). Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies 70*(2), 317–342.

Levinsohn, J. and W. Petropoulos (2001). *Creative Destruction or Just Plain Destruction? The U.S. Textile and Apparel Industries since 1972*. NBER Working Paper No. 8348.

Mairesse, J. and Z. Griliches (1990). Heterogeneity in Panel Data: Are There Stable Production Functions. In P. C. et al. (Ed.), *Essays in Honor of Edmond Malinvaud*, Volume 3, pp. 125–147. Cambridge: MIT Press.

Mairesse, J. and Z. Griliches (1998). Production Functions: The Search for Identification. In S. Strom (Ed.), *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, pp. 169–203. Cambridge University Press.

Marschak, J. and W. H. Andrews (1944). Random Simultaneous Equations and the Theory of Production. *Econometrica 12*, 143–205.

Meeusen, W. and J. van den Broeck (1977). Efficiency Estimation from Cobb-Douglas Production functions with Composed Error. *International Economic Review 8*, 435–44

Olley, G. S. and A. Pakes (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry". *Econometrica 64*(6), 1263–97.

Roberts, M. (1996). Colombia, 1977-85: Producer Turnover, Margins, and Trade Exposure. In M. Roberts and J. R. Tybout (Eds.), *Industrial Evolution in Developing Countries*, Chapter 10, pp. 227–59. New York: Oxford University Press for the World Bank.

Roberts, M. and J. Tybout (1997, September). The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs. *American Economic Review 87*, 545–64.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy 94*, 1002–37.

Schmidt, P. and R. C. Sickles (1984). Productivity Frontiers and Panel Data. *Journal of Business and Economic Statistics 2*, 367–374.

Seiford, L. L. and R. M. Thrall (1990). Recent Developments in DEA. The Mathematical Programming Approach to Frontier Analysis. *Journal of Econometrics 46*, 7–38.

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics 70*, 65–94.

Solow, R. M. (1957). Technical Change and the Aggregate Production Function. *Review of Economics and Statistics 39*, 312–320.

Stevenson, R. E. (1980). Likelihood Functions for Generalized Stochastic Frontier Estimation. *Journal of Econometrics 13*, 57–66.

Stigler, G. J. (1976, March). Xistence of X-efficiency. *American Economic Review 66*(1), 213–16.

Tybout, J. R. (2000, March). Manufacturing Firms in Developing Countries: How Well Do They Do, and Why? *Journal of Economic Literature 28*, 11–44.

Tybout, J. R. and L. Liu (1996). Productivity Growth in Chile and Colombia: The Role of Entry, Exit, and Learning. In M. Roberts and J. R. Tybout (Eds.), *Industrial Evolution in Developing Countries*, Chapter 4, pp. 73–103. New York: Oxford University Press for the World Bank.

Van Biesebroeck, J. (2003a, October). *Exporting Raises Productivity in sub-Saharan African Manufacturing Firms.* NBER Working paper No. 10020.

Van Biesebroeck, J. (2003b, January). Productivity Dynamics with Technology Choice: An Application to Automobile Assembly. *Review of Economic Studies 70*(1), 167–198.

Van Biesebroeck, J. (2003c, July). Robustness of Productivity Estimates. mimeo, University of Toronto.

Figure 1: Decomposing shifts along the frontier from a shift of the frontier

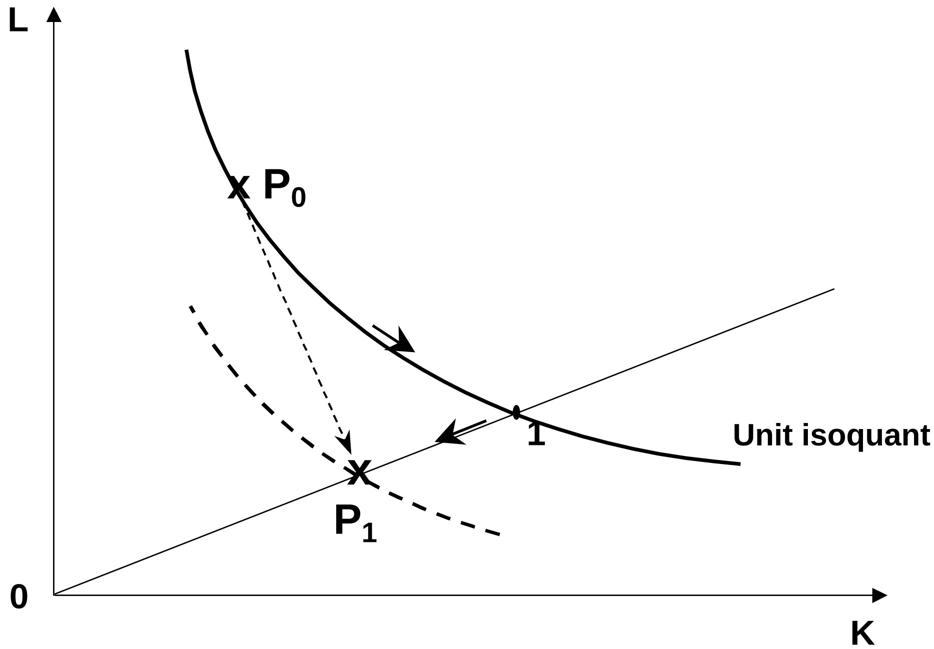

Figure 2: Nonparametric production frontiers

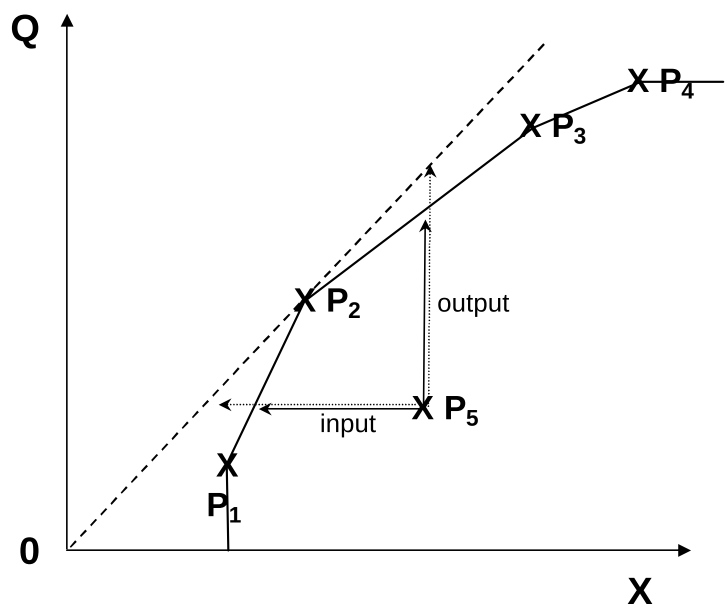

Table 1: Summary statistics for the sample of manufacturing plants in Colombia

	mean	standard deviation	5th percentile	median	95th percentile
Number of observations	14348				
Number of plants	1957				
Number of years	9.42	4.13	3	9	15
Value added	4007	30575	201	971	10511
Employment	66	141	9	30	213
Capital (in % of VA)	0.38	1.05	0.02	0.18	1.07
Investment (in % of VA)	0.09	0.31	0.00	0.01	0.38
Wage share (in % of VA)	0.55	0.22	0.20	0.54	0.86
Output growth	0.04	0.52	-0.68	0.05	0.73
Employment growth	-0.01	0.36	-0.51	0.00	0.44
Capital growth	-0.03	0.56	-0.54	-0.18	0.96

Table 2: Coefficient estimates for the production function

	CRS		VRS		
	α_l	α_k	α_l	α_k	RTS
OLS	0.83	0.17	0.89	0.17	1.06
IN[1]	0.55	0.45	0.58	0.48	1.06
DEA[2]	0.57	0.43	0.72	0.32	1.04
GMM	0.71	0.29	0.21	0.24	0.45
SF1	0.80	0.20	0.68	0.17	0.85
SF2	0.76	0.24	0.32	0.09	0.41
OP	0.75	0.25	0.74	0.18	0.92

[1] Average revenue share multiplied by RTS
[2] Average percentage weight multiplied by RTS

Table 3: Interquartile range for productivity levels

	CRS		VRS	
	25th %	75th %	25th %	75th %
OLS	-0.289	0.332	-0.279	0.324
IN	-0.384	0.415	-0.364	0.400
DEA	-0.387	0.385	-0.500	0.525
GMM	-0.318	0.350	-0.487	0.550
SF1	-0.293	0.332	-0.328	0.370
SF2	-0.287	0.313	-0.479	0.554
OP1	-0.306	0.339	-0.304	0.348
OP2	-0.047	0.050	-0.052	0.084

Table 4: Correlations between productivity level estimates

CRS \ VRS	OLS	IN	DEA	GMM	SF1	SF2	OP1	OP2
OLS	1	0.82	0.70	0.70	0.94	0.59	0.97	0.16
IN	0.81	1	0.87	0.70	0.80	0.50	0.82	-0.02
DEA	0.83	0.96	1	0.87	0.82	0.71	0.79	0.14
GMM	0.97	0.91	0.93	1	0.90	0.93	0.84	0.51
SF1	1.00	0.83	0.85	0.98	1	0.81	0.99	0.36
SF2	0.89	0.79	0.81	0.90	0.90	1	0.75	0.61
OP1	0.99	0.88	0.90	1.00	0.99	0.90	1	0.30
OP2	0.20	0.19	0.17	0.22	0.21	0.22	0.21	1

Table 5: Weighted and unweighted productivity growth averages

	CRS		VRS	
	unweighted average	weighted by aggregate input	unweighted average	weighted by aggregate input
OLS	0.060	0.029	0.061	0.026
IN	0.057	0.007	0.056	0.005
DEA	0.063	0.009	0.061	0.024
GMM	0.062	0.026	0.052	0.043
SF1	0.060	0.029	0.057	0.037
SF2	0.058	0.037	0.046	0.043
OP1	0.061	0.028	0.058	0.033
OP2	0.013	0.012	0.029	0.017

Table 6: Correlations between productivity growth estimates

CRS \ VRS	OLS	IN	DEA	GMM	SF1	SF2	OP1	OP2
OLS	1	0.90	0.83	0.90	0.99	0.50	1.00	0.03
IN	0.90	1	0.93	0.92	0.92	0.50	0.91	0.01
DEA	0.92	0.96	1	0.95	0.88	0.51	0.87	0.01
GMM	0.99	0.94	0.96	1	0.95	0.57	0.93	0.05
SF1	1.00	0.90	0.93	0.99	1	0.54	1.00	0.04
SF2	0.53	0.51	0.51	0.53	0.53	1	0.53	0.06
OP1	1.00	0.93	0.95	1.00	1.00	0.53	1	0.03
OP2	0.08	0.17	0.17	0.13	0.09	0.09	0.12	1

Table 7: Productivity premiums for exporters

	export dummy		lagged export dummy		exp. dummy	exp. share
	CRS	VRS	CRS	VRS	CRS	
OLS	0.352	0.273	0.327	0.247	0.410	-0.055
IN	0.137	0.159	0.094	0.118	0.194	-0.054
DEA	0.127	0.512	0.083	0.468	0.191	-0.061
GMM	0.273	0.989	0.239	0.978	0.332	-0.056
SF1	0.337	0.545	0.311	0.526	0.395	-0.056
SF2	0.295	1.116	0.282	1.140	0.332	-0.035
OP1	0.298	0.448	0.268	0.426	0.357	-0.056
OP2	0.080	0.206	0.093	0.229	0.084	-0.004

OLS coefficients in separate regressions of productivity levels on an export dummy and export share with time, industry, and location dummies. All coefficients are significant at the 1% level.

Table 8: Productivity growth premiums for foreign exposure dummies

	exporter	royalties paid	imported inputs	foreign workers
Fraction	0.16	0.07	0.10	0.02
OLS	0.027**	-0.023	0.011	-0.050
IN	0.001	-0.019	0.011	-0.039
DEA	0.008	-0.021	0.002	-0.058
GMM	0.036**	-0.029*	-0.003	-0.051
SF1	0.032**	-0.025	0.006	-0.050
SF2	0.019	-0.047*	-0.016	-0.087*
OP1	0.030**	-0.025	0.008	-0.050
OP2	0.006*	-0.007	-0.004	0.002

OLS coefficients of productivity growth an foreign exposure dummies (separately).
**=significant at 5% level, *=at 10% level

Table 9: Productivity growth premiums related to physical or human capital investment

	physical capital investments			human capital intensity		
	average	high	frequent	managers	high skill	technical
OLS	0.008	0.005	0.019**	0.020**	-0.011	-0.005
IN	-0.012	-0.027**	0.009	-0.006	-0.044**	0.004
DEA	0.006	-0.009	0.024**	0.001	-0.050**	-0.005
GMM	0.034**	0.022**	0.044**	0.012	-0.024**	-0.003
SF1	0.020*	0.014*	0.030**	0.019**	-0.013	-0.005
SF2	0.049**	0.023*	0.035**	0.039**	-0.025*	-0.034**
OP1	0.016*	0.010	0.026**	0.019**	-0.013	-0.005
OP2	-0.004	0.000	-0.003	0.009**	0.001	-0.004

OLS coefficients of productivity growth an investment dummies (separately).
Each dummy is equal to one for 1/3 of all plants.
**=significant at 5% level, *=at 10% level

Table 10: Productivity growth decompositions using aggregate input weights (1977-1987)

	Total growth	Within	Between	Covariance	Entry	- Exit
OLS	0.384	0.290	-0.016	-0.025	0.158	-0.023
DEA	0.456	0.298	0.014	-0.071	0.225	-0.008
GMM	0.419	0.296	-0.006	-0.029	0.185	-0.027
SF1	0.391	0.291	-0.015	-0.025	0.163	-0.024
SF2	0.413	0.261	-0.023	-0.015	0.185	0.004
OP1	0.408	0.294	-0.010	-0.026	0.176	-0.026
OP2	-0.118	-0.039	0.015	0.036	-0.085	-0.016
LP	0.324	0.289	-0.022	-0.046	0.116	-0.014

LP is the logarithm of labor productivity (value added over employment).

Table 11: Productivity growth decompositions using output weights (1977-1987)

	Total growth	Within	Between	Covariance	Entry	- Exit
OLS	0.323	0.217	-0.055	0.103	0.172	-0.115
DEA	0.387	0.175	-0.060	0.108	0.198	-0.034
GMM	0.355	0.212	-0.057	0.106	0.191	-0.096
SF1	0.329	0.216	-0.055	0.104	0.176	-0.111
SF2	0.368	0.198	-0.061	0.105	0.217	-0.090
OP1	0.345	0.214	-0.056	0.105	0.185	-0.102
OP2	-0.170	-0.027	-0.011	0.026	-0.092	-0.066
LP	0.275	0.224	-0.052	0.100	0.145	-0.141

LP is the logarithm of labor productivity (value added over employment).